

OIFE

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/954,846

DATE: 01/08/2002

TIME: 10:07:23

Input Set : N:\Crf3\RULE60\09954846.raw

Output Set: N:\CRF3\01082002\I954846.raw

SEQUENCE LISTING

3 (1) GENERAL INFORMATION:

5 (i) APPLICANT: Tang, Y. Tom
 6 Corley, Neil C.
 7 Guegler, Karl J.
 8 Patterson, Chandra
 9 Baughn, Mariah R.

ENTERED

11 (ii) TITLE OF INVENTION: THIOREDOXIN PROTEINS

13 (iii) NUMBER OF SEQUENCES: 15

15 (iv) CORRESPONDENCE ADDRESS:

16 (A) ADDRESSEE: Incyte Pharmaceuticals, Inc.
 17 (B) STREET: 3174 Porter Drive
 18 (C) CITY: Palo Alto
 19 (D) STATE: CA
 20 (E) COUNTRY: USA
 21 (F) ZIP: 94304

23 (v) COMPUTER READABLE FORM:

24 (A) MEDIUM TYPE: Diskette
 25 (B) COMPUTER: IBM Compatible
 26 (C) OPERATING SYSTEM: Windows
 27 (D) SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

29 (vi) CURRENT APPLICATION DATA:

C--> 30 (A) APPLICATION NUMBER: US/09/954,846
 C--> 31 (B) FILING DATE: 17-Sep-2001
 32 (C) CLASSIFICATION:

34 (vii) PRIOR APPLICATION DATA:

35 (A) APPLICATION NUMBER: 09/107,248
 36 (B) FILING DATE:

38 (viii) ATTORNEY/AGENT INFORMATION:

39 (A) NAME: Cerrone, Michael C
 40 (B) REGISTRATION NUMBER: 39,132
 41 (C) REFERENCE/DOCKET NUMBER: PF-0556 US

43 (ix) TELECOMMUNICATION INFORMATION:

44 (A) TELEPHONE: 650-855-0555
 45 (B) TELEFAX: 650-855-0572
 46 (C) TELEX:

48 (2) INFORMATION FOR SEQ ID NO: 1:

50 (i) SEQUENCE CHARACTERISTICS:

51 (A) LENGTH: 172 amino acids
 52 (B) TYPE: amino acid
 53 (C) STRANDEDNESS: single
 54 (D) TOPOLOGY: linear

56 (vii) IMMEDIATE SOURCE:

57 (A) LIBRARY: BRSTNOT02
 58 (B) CLONE: 1925679

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1 :

62 Met Glu Thr Arg Pro Arg Leu Gly Ala Thr Cys Leu Leu Gly Phe

RAW SEQUENCE LISTING

DATE: 01/08/2002

PATENT APPLICATION: US/09/954,846

TIME: 10:07:23

Input Set : N:\Crf3\RULE60\09954846.raw

Output Set: N:\CRF3\01082002\I954846.raw

63	5	10	15
64	Ser Phe Leu Leu Leu Val Ile Ser Ser Asp Gly His Asn Gly Leu		
65	20	25	30
66	Gly Lys Gly Phe Gly Asp His Ile His Trp Arg Thr Leu Glu Asp		
67	35	40	45
68	Gly Lys Lys Glu Ala Ala Ala Ser Gly Leu Pro Leu Met Val Ile		
69	50	55	60
70	Ile His Lys Ser Trp Cys Gly Ala Cys Lys Ala Leu Lys Pro Lys		
71	65	70	75
72	Phe Ala Glu Ser Thr Glu Ile Ser Glu Leu Ser His Asn Phe Val		
73	80	85	90
74	Met Val Asn Leu Glu Asp Glu Glu Glu Pro Lys Asp Glu Asp Phe		
75	95	100	105
76	Ser Pro Asp Gly Gly Tyr Ile Pro Arg Ile Leu Phe Leu Asp Pro		
77	110	115	120
78	Ser Gly Lys Val His Pro Glu Ile Ile Asn Glu Asn Gly Asn Pro		
79	125	130	135
80	Ser Tyr Lys Tyr Phe Tyr Val Ser Ala Glu Gln Val Val Gln Gly		
81	140	145	150
82	Met Lys Glu Ala Gln Glu Arg Leu Thr Gly Asp Ala Phe Arg Lys		
83	155	160	165
84	Lys His Leu Glu Asp Glu Leu		
85	170		

90 (2) INFORMATION FOR SEQ ID NO: 2:

92 (i) SEQUENCE CHARACTERISTICS:

93 (A) LENGTH: 258 amino acids

94 (B) TYPE: amino acid

95 (C) STRANDEDNESS: single

96 (D) TOPOLOGY: linear

98 (vii) IMMEDIATE SOURCE:

99 (A) LIBRARY: BRAINOT19

100 (B) CLONE: 3244141

102 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2 :

104	Met Ala Val Leu Ala Pro Leu Ile Ala Leu Val Tyr Ser Val Pro		
105	5	10	15
106	Arg Leu Ser Arg Trp Leu Ala Gln Pro Tyr Tyr Leu Leu Ser Ala		
107	20	25	30
108	Leu Leu Ser Ala Ala Phe Leu Leu Val Arg Lys Leu Pro Pro Leu		
109	35	40	45
110	Cys His Gly Leu Pro Thr Gln Arg Glu Asp Gly Asn Pro Cys Asp		
111	50	55	60
112	Phe Asp Trp Arg Glu Val Glu Ile Leu Met Phe Leu Ser Ala Ile		
113	65	70	75
114	Val Met Met Lys Asn Arg Arg Ser Met Phe Leu Met Thr Cys Lys		
115	80	85	90
116	Pro Pro Leu Tyr Met Gly Pro Glu Tyr Ile Lys Tyr Phe Asn Asp		
117	95	100	105
118	Lys Thr Ile Asp Glu Glu Leu Glu Arg Asp Lys Arg Val Thr Trp		
119	110	115	120

RAW SEQUENCE LISTING

DATE: 01/08/2002

PATENT APPLICATION: US/09/954,846

TIME: 10:07:23

Input Set : N:\Crif3\RULE60\09954846.raw

Output Set: N:\CRF3\01082002\I954846.raw

```

120 Ile Val Glu Phe Phe Ala Asn Trp Ser Asn Asp Cys Gln Ser Phe
121                               125                               130                               135
122 Ala Pro Ile Tyr Ala Asp Leu Ser Leu Lys Tyr Asn Cys Thr Gly
123                               140                               145                               150
124 Leu Asn Phe Gly Lys Val Asp Val Gly Arg Tyr Thr Asp Val Ser
125                               155                               160                               165
126 Thr Arg Tyr Lys Val Ser Thr Ser Pro Leu Thr Lys Gln Leu Pro
127                               170                               175                               180
128 Thr Leu Ile Leu Phe Gln Gly Gly Lys Glu Ala Met Arg Arg Pro
129                               185                               190                               195
130 Gln Ile Asp Lys Lys Gly Arg Ala Val Ser Trp Thr Phe Ser Glu
131                               200                               205                               210
132 Glu Asn Val Ile Arg Glu Phe Asn Leu Asn Glu Leu Tyr Gln Arg
133                               215                               220                               225
134 Ala Lys Lys Leu Ser Lys Ala Gly Asp Asn Ile Pro Glu Glu Gln
135                               230                               235                               240
136 Pro Val Ala Ser Thr Pro Thr Thr Val Ser Asp Gly Glu Asn Lys
137                               245                               250                               255

```

138 Lys Asp Lys

143 (2) INFORMATION FOR SEQ ID NO: 3:

145 (i) SEQUENCE CHARACTERISTICS:

146 (A) LENGTH: 1440 base pairs

147 (B) TYPE: nucleic acid

148 (C) STRANDEDNESS: single

149 (D) TOPOLOGY: linear

151 (vii) IMMEDIATE SOURCE:

152 (A) LIBRARY: BRSTNOT02

153 (B) CLONE: 1925679

155 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3 :

```

157 TCTGGGAAGT AGAGGTGTTG TGCTGAGCGG CGCTCGGCGA ACTGTGTGGA CCGTCTGCTG 60
158 GGAATCCGGC CCTGCGTCCG CTCAGCCCCG TGGCCCCGCG CACCTACTGC CATGGAGACG 120
159 CGGCCTCGTC TCGGGGCCAC CTGTTTGCTG GGCTTCAGTT TCCTGCTCCT CGTCATCTCT 180
160 TCTGATGGAC ATAATGGGCT TGGAAAGGGT TTTGGAGATC ATATTCATTG GAGGACACTG 240
161 GAAGATGGGA AGAAAGAAGC AGCTGCCAGT GGACTGCCCC TGATGGTGAT TATTCATAAA 300
162 TCCTGGTGTG GAGCTTGCAA AGCTCTAAAG CCCAAATTTG CAGAATCTAC GGAAATTTCA 360
163 GAACTCTCCC ATAATTTTGT TATGGTAAAT CTTGAGGATG AAGAGGAACC CAAAGATGAA 420
164 GATTTTCAGC CTGACGGGGG TTATATTCCA CGAATCCTTT TTCTGGATCC CAGTGGCAAG 480
165 GTGCATCCTG AAATCATCAA TGAGAATGGA AACCCAGCT ACAAGTATTT TTATGTCAGT 540
166 GCCGAGCAAG TTGTTTCAGG GATGAAGGAA GCTCAGGAAA GGCTGACGGG TGATGCCTTC 600
167 AGAAAGAAAC ATCTTGAAGA TGAATTGTAA CATGAATGTG CCCCTTCTTT CATCAGAGTT 660
168 AGTGTCTCTG AAGGAAAGCA GCAGGGAAGG GAATATTGAG GAATCATCTA GAACAATTAA 720
169 GCCGACCAGG AAACCTCATT CCTACCTACA CTGGAAGGAG CGCTCTCACT GTGGAAGAGT 780
170 TCTGCTAACA GAAGCTGGTC TGCATGTTTG TGGATCCAGC GGAGAGTGGC AGACTTTCTT 840
171 CTCCTTTTCC CTCTCACCTA AATGTCAACT TGTCATTGAA TGTAAGAAT GAAACCTTCT 900
172 GACACAAAAC TTGAGCCACT TGGATGTTTA CTCCTCGCAC TTAAGTATTT GAGTCTTTTC 960
173 CCATTTCTCT CCACTTTACT CACCTTAGTG GTGAAAGGAG ACTAGTAGCA TCTTTTCTAC 1020
174 AACGTTAAAA TTGCAGAAGT AGCTTATCAT TAAAAACAA CAACAACAAC AATAACAATA 1080
175 AATCCTAAGT GTAAATCAGT TATTCTACCC CCTACCAAGG ATATCAGCCT GTTTTTTCCC 1140
176 TTTTTTCTCC TGGGAATAAT TGTGGGCTTC TTCCCAAATT TCTACAGCCT CTTTCTCTT 1200

```

RAW SEQUENCE LISTING

DATE: 01/08/2002

PATENT APPLICATION: US/09/954,846

TIME: 10:07:23

Input Set : N:\Crif3\RULE60\09954846.raw

Output Set: N:\CRF3\01082002\I954846.raw

177 CTCATGCTTG AGCTTCCCTG TTTGCACGCA TGCCTGTGCA GGACTGGCTG TGTGCTTGGA 1260
 178 CTCGGCTCCA GGTGGAAGCA TGCTTTCCCT TGTTACTGTT GGAGAACTC AAACCTTCAA 1320
 179 GCCCTAGGTG TAGCCATTTT GTCAAGTCAT CAACTGTATT TTTGTACTGG CATTAACAAA 1380
 180 AAAAGAGATA AAATATTGTA CCATTAAACT TTAATAAAAC TTTAAAAGGA AAAAAAAAAA 1440

183 (2) INFORMATION FOR SEQ ID NO: 4:

185 (i) SEQUENCE CHARACTERISTICS:

186 (A) LENGTH: 1555 base pairs

187 (B) TYPE: nucleic acid

188 (C) STRANDEDNESS: single

189 (D) TOPOLOGY: linear

191 (vii) IMMEDIATE SOURCE:

192 (A) LIBRARY: BRAINOT19

193 (B) CLONE: 3244141

195 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4 :

197 AGGGGAGGCG GGGCGAGACC TACGACGCCG GCGAGCAGTG GCCGTTACGG CCGAAAAGAT 60
 198 GGC GGCTCTTG GCACCTCTAA TTGCTCTCGT GTATTTCGGTG CCGCGACTTT CACGATGGCT 120
 199 CGCCCAACCT TACTACCTTC TGTCGGCCCT GCTCTCTGCT GCCTTCCTAC TCGTGAGGAA 180
 200 ACTGCCGCCG CTCTGCCACG GTCTGCCAC CCAACGCGAA GACGGTAACC CGTGTGACTT 240
 201 TGA CTGGAGA GAAGTGAGGA TCCTGATGTT TCTCAGTGCC ATTGTGATGA TGAAGAACCG 300
 202 CAGATCCATG TTCCTGATGA CGTGCAAAACC CCCCTATAT ATGGGCCCTG AGTATATCAA 360
 203 GTACTTCAAT GATAAAACCA TTGATGAGGA ACTAGAACGG GACAAGAGGG TCACTTGGAT 420
 204 TGTGGAGTTC TTTGCCAATT GGTCTAATGA CTGCCAATCA TTTGCCCTA TCTATGCTGA 480
 205 CCTCTCCCTT AAATACAAC GTACAGGGCT AAATTTTGGG AAGGTGGATG TTGGACGCTA 540
 206 TACTGATGTT AGTACGCGGT ACAAGTGAG CACATCACCC CTCACCAAGC AACTCCCTAC 600
 207 C CTGATCCTG TTCCAAGGTG GCAAGGAGGC AATGCGGCGG CCACAGATTG ACAAGAAAGG 660
 208 AC GGCTGTC TCATGGACCT TCTCTGAGGA GAATGTGATC CGAGAATTTA ACTTAAATGA 720
 209 GCTATACCAG CGGGCCAAGA AACTATCAAA GGCTGGAGAC AATATCCCTG AGGAGCAGCC 780
 210 TGTGGCTTCA ACCCCCACCA CAGTGTGAGA TGGGGAAAAC AAGAAGGATA AATAAGATCC 840
 211 TCACTTTGGC AGTGCTTCCT CTCCTGTCAA TTCCAGGCTC TTTCCATAAC CACAAGCCTG 900
 212 AGGCTGCAGC CTTTTATTTA TGTTTCCCT TTGGCTGTGA CTGGGTGGGG CAGCATGCAG 960
 213 CTTCTGATTT TAAAGAGGCA TCTAGGGAAT TGTCAGGCAC CCTACAGGAA GGCCTGCCAT 1020
 214 GCTGTGGCCA ACTGTTTCAC TGGAGCAAGA AAGAGATCTC ATAGGACGGA GGGGGAAATG 1080
 215 GTTTCCCTCC AAGCTTGGGT CAGTGTGTTA ACTGCTTATC AGCTATTCAG ACATCTCCAT 1140
 216 GGTCTCTCCA TGAAACTCTG TGGTTTCATC ATTCCTTCTT AGTTGACCTG CACAGCTTGG 1200
 217 TTAGACCTAG ATTTAACCCT AAGGTAAGAT GCTGGGGTAT AGAACGCTAA GAATTTTCCC 1260
 218 CCAAGGACTC TTGCTTCCTT AAGCCCTTCT GGCTTCGTTT ATGGTCTTCA TTTAAAGTAT 1320
 219 AAGCCTAACT TTGTCGCTAG TCCTAAGGAG AAACCTTTAA CCACAAAGTT TTTATCATTTG 1380
 220 AAGACAATAT TGAACAACCC CCTATTTTGT GGGGATTGAG AAGGGGTGAA TAGAGGCTTG 1440
 221 AGACTTTCCT TTGTGTGGTA GGACTTGGAG GAGAAATCCC CTGGACTTTC ACTAACCTTC 1500
 222 TGACATACTC CCCACACCCA GTTGATGGCT TTCCGTAATA AAAAGATTGG GATTA 1555

225 (2) INFORMATION FOR SEQ ID NO: 5:

227 (i) SEQUENCE CHARACTERISTICS:

228 (A) LENGTH: 212 base pairs

229 (B) TYPE: nucleic acid

230 (C) STRANDEDNESS: single

231 (D) TOPOLOGY: linear

233 (vii) IMMEDIATE SOURCE:

234 (A) LIBRARY: BRSTNOT02

235 (B) CLONE: 1925679H1

RAW SEQUENCE LISTING

DATE: 01/08/2002

PATENT APPLICATION: US/09/954,846

TIME: 10:07:23

Input Set : N:\Cr3\RULE60\09954846.raw

Output Set: N:\CRF3\01082002\I954846.raw

```

237      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5 :
239 GATGGGAAGA AAGAAGCAGC TGCCAGTGGA CTGCCCCTGA TGGTGATTAT TCATAAATCC 60
240 TGGTGTGGAG CTTGCAAAGC TCTAAAGCCC AAATTTGCAG AATCTACGGA AATTTTCAGAA 120
241 CTCTCCCATATA ATTTTGTAT GGTAAATCTT GAGGATGAAG AGGAACCCAA AGATGAAGAT 180
242 TTCAGCCCTG ACGGGGGTTA TATTCCACGA AT 212
245 (2) INFORMATION FOR SEQ ID NO: 6:
247      (i) SEQUENCE CHARACTERISTICS:
248          (A) LENGTH: 248 base pairs
249          (B) TYPE: nucleic acid
250          (C) STRANDEDNESS: single
251          (D) TOPOLOGY: linear
253      (vii) IMMEDIATE SOURCE:
254          (A) LIBRARY: ENDANOT01
255          (B) CLONE: 2456812H1
257      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6 :
259 ATCTGGGAAG TAGAGGTGTT GTGCTGAGCG GCGCTCGGCG AACTGTGTGG ACCGTCTGCT 60
260 GGGACTCCGG CCTTGCGTCC GCTCAGCCCC GTGGCCCCGC GCACCTACTG CCATGGAGAC 120
261 GCGGCCTCGT CTCGGGGCCA CCTGTTTGCT GGGCTTCAGT TTCCTGCTCC TCGTCATCTC 180
262 TTCTGATGGA CATAATGGGC TTGGAAGGG TTTTGGAGAT CATATTCATT GGAGGACACT 240
263 GGAAGATG 248
266 (2) INFORMATION FOR SEQ ID NO: 7:
268      (i) SEQUENCE CHARACTERISTICS:
269          (A) LENGTH: 541 base pairs
270          (B) TYPE: nucleic acid
271          (C) STRANDEDNESS: single
272          (D) TOPOLOGY: linear
274      (vii) IMMEDIATE SOURCE:
275          (A) LIBRARY: BRSTNOT02
276          (B) CLONE: 1925679R6
278      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7 :
280 GATGGGAAGA AAGAAGCAGC TGCCAGTGGA CTGCCCCTGA TGGTGATTAT TCATAAATCC 60
281 TGGTGTGGAG CTTGCAAAGC TCTAAAGCCC AAATTTGCAG AATCTACGGA AATTTTCAGAA 120
282 CTCTCCCATATA ATTTTGTAT GGTAAATCTT GAGGATGAAG AGGAACCCAA AGATGAAGAT 180
283 TTCAGCCCTG ACGGGGGTTA TATTCCACGA ATCCTTTTTC TGGATCCCAG TGGCAAGGTG 240
284 CATCCTGAAA TCATCAATGA GAATGGAAC CCCAGCTACA AGTATTTTTA TGTCAGTGCC 300
285 GAGCAAGTTG TTCAGGGGAT GAAGGAAGCT CAGGAAAGGC TGACGGGTGA TGCCTTCAGA 360
286 AAGAAACATC TTGAAGATGA ATTGTAACAT GAATGTGCCC CTTCTTTCAT CAGAGTTAGT 420
287 GTTCTGGAAG GAAAGCAGCA GGGAAGGGAA TATTGAGGAA TCATCTAGAA CAATTAAGCC 480
288 GACCAGGAAA CCTCATTCCT ANCTACACTG GAANGAGCGC TCTCACTGTG GAAGAGTTCT 540
289 G 541
292 (2) INFORMATION FOR SEQ ID NO: 8:
294      (i) SEQUENCE CHARACTERISTICS:
295          (A) LENGTH: 578 base pairs
296          (B) TYPE: nucleic acid
297          (C) STRANDEDNESS: single
298          (D) TOPOLOGY: linear
300      (vii) IMMEDIATE SOURCE:
301          (A) LIBRARY: BLADTUT04
302          (B) CLONE: 1522838F1

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/954,846

DATE: 01/08/2002

TIME: 10:07:24

Input Set : N:\Crf3\RULE60\09954846.raw

Output Set: N:\CRF3\01082002\I954846.raw

L:30 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]

L:31 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]

L:337 M:111 C: (47) String data converted to upper case,

M:111 Repeated in SeqNo=9